THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Constance Franklin Kenney

TITLE OF THESIS Estimating the Reliability of the Short Forms of the WISC-R

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Education

YEAR THIS DEGREE GRANTED 1977

THE UNIVERSITY OF ALBERTA

ESTIMATING THE RELIABILITY OF THE SHORT FORMS OF THE WISC-R

by

© CONSTANCE FRANKLIN KENNEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Estimating the Reliability of the Short Forms of the WISC-R submitted by Constance Franklin Kenney in partial fulfilment of the requirements for the degree of Master of Education.

ABSTRACT

This study was undertaken to fulfill a need for more information concerning the reliabilities of the short forms of the WISC-R when used in a psycho-educational clinic. Before this assessment could be done a preliminary study, requiring the analysis of the standardization data of the WISC-R, was required to determine which short forms were most valid. The resulting 40 best short forms were tested for reliability on the clinic sample.

The concept of intelligence was discussed as well as the reliability and validity of the WISC-R. An outline of some of the research using a variety of the short forms was given. The relative merits and limitations of the "split-half" and "subtest combination" short forms were evaluated. It was concluded that the subtest combination short forms only had merit statistically.

The literature reviewed explained the spuriously high component in correlations involving two sets of scores obtained in just one administration of a test, and the process that was necessary to resolve the problem, namely, a corrected or revised formula for use with part-whole relationships.

Data from the standardization sample which were all available in the manual, were used to obtain a list of 40 best short forms after which their reliabilities with the research sample were analysed. In addition, comparisons between the short forms of the WISC and WISC-R were made and their similarities and differences discussed.

The results of the study were clear and unequivocal in their

support for the use of selected short forms of the WISC-R with individuals who might be referred to a psycho-educational clinic or school counsellor. Some suggestions on how to select the appropriate short forms were made.

The study contributed to the accumulating body of research supporting the reputation of the WISC-R as an even more robust test than its predecessor.

ACKNOWLEDGEMENTS

My sincere gratitude is offered to the following persons:

Dr. Henry Janzen, who gave so freely of his ideas and time, for his interest, encouragement and support.

Dr. Len Stewin, for his kind consideration and constructive criticism.

Professor Jeanette Funke, for her friendship, interest and enthusiasm.

Dr. Steve Hunka and Dr. Tom McGuire, for their expert guidance and assistance in the analysis of my data.

Mrs. Margaret Voice for her exceptional skills and concern in typing this thesis.

And very special thanks to my husband, Ted Kenney, for his continuous support, patience and understanding.

And finally, to my mother, Mollee Franklin, who very early in my life gave me the belief in myself which enabled me to prevail.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION TO THE STUDY

The Problem

Every year thousands of school children and adolescents are administered psychological tests in order to help teachers, parents and clinicians make important decisions regarding the development and progress of each individual child. Such tests are most frequently administered to children who have demonstrated that they are having some difficulty in adjusting to the school situation. Either they are not progressing at the expected rate in the basic school subjects or they are experiencing emotional and/or behavioral problems which appear to be interfering with their learning progress.

Foremost in any selection of tests is an intelligence test. If at all feasible an individually administered test such as the Stanford-Binet or one of the Wechsler Scales is used in preference to a group administered test such as the Lorge-Thorndike or Otis-Lennon. Although more expensive and time-consuming to administer, the individual tests offer a much more dependable measure of the child's intellectual ability. In addition, it offers an opportunity for the examiner to evaluate behavior in the test situation which might relate to performance in the classroom, such as the child's ability to understand what is expected of him, his attitude, his method of attacking the question and in general, any behaviors which might be hindering or interfering with his best performance.

Although the most reliable and valid of all tests, an intelligence test alone is rarely sufficient to diagnose a child's learning difficulties. In addition it is usually necessary to administer other kinds of tests in order to evaluate other aspects of his behavior. For example, examiners are interested in knowing: At what level is this child reading? What are his spelling and arithmetic skills? Does he speak well and get along well with his peers? Does he have perceptual-motor difficulties? Are aural and visual acuity adequate? For this reason, most examiners would agree that when a child is referred for assessment several tests rather than a single intelligence test is the preferred procedure.

While a carefully selected and properly administered battery of tests can offer a wealth of information, in most situations it is more an ideal than a reality. Such programs are found to be prohibitively expensive and very time consuming. The psychologist's or counsellor's dilemma has always been, and continues to be, how to get the most useful information concerning a child while making the best use of the time and resources available to him.

Background to the Problem

In response to the question of how to make the best use of time and resources available the first line of attack has been the development of better tests and a more thorough understanding and utilization of the test results. In the area of intelligence testing psychologists have explored two avenues of research in an effort to obtain the most information in as short a time as possible. One direction has led to the development and use of various abbreviated or modified short forms

of the widely used Wechsler Scales (Doppelt, 1956; Silverstein, 1970; Tellegen & Briggs, 1967; Yudin, 1966), while the other direction has focused on a scatter or profile analysis of the individual subtest functions within the full-length test (Frost & Frost, 1962; Rhodes, 1969; Schofield, 1952).

The rationale of the short forms is that because of their high correlation with the full-length test they provide a reliable measure of global intelligence, that is, they are a good measure of an individual's overall capacity for understanding and coping with his world, while meeting an increasing demand for brief psychological tests (Clements, 1965; Patterson, 1953; Silverstein, 1970; Wechsler, 1958). On the other hand, the profile or scatter analysis approach to intelligence testing derives its rationale from ego psychology (Fromm et al, 1957; Rapaport et al, 1968). Variation among subtests is considered a means for assessing such ego functions as memory, concentration, attention, judgment, planning and concept formation (Rapaport et al, 1968). An evaluation of different subtests provides indices for development in specific areas and allows inferences to be made about personality, thought processes, defences, et cetera (Frost & Frost, 1962; Blatt & Allison, 1968).

In summary, the goal in using a short form is to get a quick and reliable estimate of global intelligence whereas the goal in using scatter analysis is to search for as much specific information as possible on each of the separate abilities or subtests of the test in order to gain as broad an understanding as possible from the administration of one high quality test. It almost goes without saying that

the latter approach, in addition to requiring more administration time also requires a high degree of experience and interpretive ability on the part of the examiner. The primary reason for using a short form is that, theoretically, it allows more time for other supplemental kinds of testing which are necessary for a complete assessment. This paper is concerned with use of the various short forms of the WISC-R as a sound and sensible method of intelligence testing.

Short Forms

Historically, the Wechsler short forms have been generally of two kinds, namely, the "subtest combination" short form and the "split-half" short form.

In the subtest combination short form only certain selected subtests are administered. From the scores on these subtests, a full scale score and IQ is estimated. For example, the Vocabulary and Block Design subtests together are considered to be a suitable short form. Studies show that scores on these two subtests correlate very highly with the full scale score, although the standard error is somewhat higher (Kaufman, 1975; Silverstein, 1967a, 1970a). Combinations of two, three, four, five and even one single subtest have been considered as suitable short forms because of their high predictive validity. As a result, short forms have been recommended for use in lieu of the full-length test in certain circumstances.

Although a single subtest, such as Vocabulary can serve as an adequate short form (Kaufman, 1975) the standard error is quite high so that while it yields a reasonably accurate estimate of IQ, the

confidence one can place on the estimate is less than with a longer test. With all short forms, the standard error of measurement is systematically reduced as the number of subtests in the short form is increased so that the confidence level is greatest for the longest short form. It can always be assumed that the predictive validity will increase, although sometimes only slightly, as the length of the short form is increased with the most accurate prediction being based on the results of the full-length test. These are some of the facts which have to be taken into consideration by the psychologist or clinician when deciding whether or not a short form could be used and if so, which one.

The "split-half" short form gets its name from the statistical procedure which reduces the length of the test by using only half the items. While the length of the test is reduced by half, a sampling from each subtest is retained. For example, by selecting every odd item in the Arithmetic subtest, every even item in the Vocabulary subtest, et cetera, a suitably sampled short form is developed. A few examples of this kind of short form (Yudin, 1967; Silverstein, 1968) are given in Appendix A. Because no subtest is omitted, a measure of all areas of intellectual functioning is obtained. For a short time after it first appeared the split-half procedure was eagerly received by clinicians because it lent itself so well to profile analysis. Later, it came under serious scrutiny for its statistical claims in the area of test reliability and following that its predictive validity. The development of these short forms will be discussed more fully in the review of the literature.

Rationale for Investigation

The thesis of this study is that the subtest combination short forms of the WISC and subsequently the WISC-R are a very practical and highly valid measure of intelligence. While these short forms have been given much attention and consideration in research and statistical studies, they have not been widely adopted for clinical purposes, mostly it would seem because of doubts associated with the ability of the short form to be a viable alternative to the full-length test.

Recent research (Kaufman, 1975; Sattler, 1974; Silverstein, 1967, 1971, 1974) has unhesitatingly supported the high validity of several of the short forms and recommended their use in testing as a general screening device and as a fast, accurate and economical estimate of intellectual ability. On the other hand, it has never been suggested and this study does not suggest, that a short form could replace the full-length test when a comprehensive assessment of intellectual functioning was required, nor would it be recommended for use when making significant classification decisions concerning mental retardation or special education (Finch, Ollendick & Ginn, 1973).

Statement of the Problem

The present study is designed to estimate the reliabilities of a select group of 40 subtest combination short forms of the WISC-R using 182 protocols obtained in a psycho-educational clinic. The purpose will be to determine whether or not any of these short forms are suitable for use in such a specialized clinic.

The best short forms will be selected by computing the

validities for all possible subtest combinations of two, three, four and five, using the WISC-R standardization sample data presented in the WISC-R manual (1974) and following the procedure outlined by Silverstein (1970) in his evaluation of the best short forms for the WISC. The selected short forms will be the top ten of each combination. The result will be a list of 40 suitable short forms from which any examiner may select the one most appropriate for his purpose.

The concepts of reliability and validity are used throughout this study to describe characteristics of the short form and its relationship to the full-length test. To clarify the meaning of these terms in the context of the study it should be mentioned that when scores on a short form are correlated with those on the full-length test, the result is considered to be a measure of the validity of the short form, that is, a measure of the degree to which the short form is measuring the same thing as the full-length test. In the studies using the standardization data to determine the 40 best short forms, for example, Silverstein (1970), Sattler (1974), and Part 1 of the present study, the resulting coefficients are a measure of validity.

When the short forms are to be used in lieu of the full-length test, as is the case in Part 2 of this study, what is required is an understanding of the dependability of the short form. How reliable is the short form? Does it consistently give an accurate estimate of IQ? How does the reliability of the short form compare to that of the full-length test? The cross validation analysis is a measure of the reliability of the short forms.

Significance of the Study

The decision to examine the reliability of the short form of the WISC-R was made for several reasons. Because intelligence tests and particularly the WISC-R, are so widely used in schools and clinics, it was felt that the short form could not be ignored by serious users of this test, particularly after recent studies which show very high validities for certain of these forms (Silverstein, 1970; Kaufman, 1975). It is important to know which, if any, of these short forms can be considered a viable alternative for assessments done by Psycho-Educational Clinics.

If the reliability and validity of the short forms for this specialized sample are as high as some studies indicate they might be, then their advantages would be well demonstrated. Not only would these tests be worthwhile because of time saved in administering the tests but the administration skills necessary to administer the short form would be less demanding and therefore more easily acquired which, clearly, would be an important advantage. Teachers themselves might well be able to do some of this testing on their own.

In addition, if short forms are a serious alternative for this particular sample, emphasis could move toward more criterion-related testing in addition to the IQ measure and perhaps more follow-up testing and evaluation of recommendations.

Finally, it was felt that the study provided an opportunity to build a data base from existing files in the Education Clinic and to commence research with clinical data which might lead to other research questions.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The Wechsler Intelligence Scale for Children, Revised (WISC-R) is the 1974 revision of the WISC which was first published in 1949 and which has proven by research and reputation to be one of the most rigorous and highly regarded intelligence tests in use today. Changes made in the WISC-R were primarily a rewording of some of the more outdated items, a modernization of subtest stimuli, some clarification for test administrators, and an adjustment in the age range to make it more compatible with the Wechsler test for younger children (WPPSI) and the one for adults (WAIS). However, the WISC-R is substantially the same as the WISC; that is, in all essential areas, the WISC-R is viewed as equivalent to the older test.

As more and more comparative research is being done on the two tests, it is gradually being conceded that in virtually all cases, the information acquired from research on the WISC can be generalized to the WISC-R with, in some cases, a tendency towards slightly more positive results on the WISC-R. This trend has led researchers to conclude that the WISC-R is an even finer test than its predecessor (Wechsler, 1974; Sattler, 1974; Kaufman, 1975; Silverstein, 1974). For this reason, in the following discussion it is quite reasonable to assume that statements and conclusions made concerning the WISC, also apply to the WISC-R.

The Concept of Intelligence as Expressed by the WISC-R

Wechsler (1974) describes intelligence as the overall capacity of a person to understand and cope with his world; he refers to it as a global entity meaning it is something multidetermined and multifaceted. The WISC-R supports this concept of intelligence by using twelve different subtests to measure as many different intellectual abilities with each contributing to the quotient considered to be a measure of general intelligence. No selected abilities are considered to be more crucial or important than the others.

The concept of mental age (MA) whereby a person was said to be functioning at a certain age level of mentality is not used with the WISC-R. Instead of an age-equivalent index, the deviation IQ, which is a measure of an individual's mental functioning in relation to others his own age, is used. The deviation IQ indicates the child's relative position in the age group to which he belongs. Therefore, a child's IQ will not vary from year to year unless his actual performance, compared to his peers, has changed.

Intelligence as measured by the WISC-R does not include all intellectual abilities but it does include those abilities which are considered valuable in our culture. For this reason, primary abilities such as conceptual learning, abstract reasoning and abstract or verbal problem-solving are emphasized in the WISC-R as they are in most intelligence tests (Wechsler, 1958). Intelligence as determined by the WISC-R is related to socially valued criteria; it is a measure of the degree to which an individual has those intellectual qualities which will get him through school and college and into the higher

socio-economic occupations (Jensen, 1967). Biases inherent in an intelligence test favor convergent, analytical and scientific thinking while, at the same time almost ignoring the divergent, imaginative and artistic processes; it will always favor the scientists over the artists (Hudson, 1971).

Reliability and Validity of the WISC-R

The reliability coefficients for the WISC-R subtests and Verbal, Performance and Full Scale IQs are reported in the manual for 11 age groups (6 1/2 through to 16 1/2 years by one year intervals) in addition to an averaged index (Wechsler, 1974). The most reliable subtests were Vocabulary (.86), Information (.85), Block Design (.85) and Similarities (.81); the lowest reliabilities were for Object Assembly, Coding, Mazes and Picture Arrangement (.70-.73). Subtest correlations were split-half correlations, except for Coding and Digit Span which were obtained by a test-retest analysis. Reliability for the IQ scales were determined by a formula for a composite group. Reliability coefficients for Verbal IQ ranged from .91 to .96 (average .94), for Performance IQ from .89 to .91 (average .90) and for Full Scale IQ from .95 to .96 (average .96).

Stability coefficients for three age groups, 6 1/2-7 1/2, 10 1/2-11 1/2 and 14 1/2-15 1/2 (N = 303) were reported (Wechsler, 1974). They were determined by retesting the group after a one-month interval and are as follows: average coefficients for Verbal subtests ranged from .77 (Digit Span) to .88 (Information); average coefficients for Performance subtests ranged from .65 (Mazes) to .81 (Block Design); average coefficient for Verbal IQ was .93, for Performance IQ .90 and

for Full Scale IQ .95.

Congruent validity for the WISC-R has been determined by correlating WISC-R scores with those on the WAIS and WPPSI for the age groups which overlap (age 6 1/2 and 16 1/2). Correlations with the WAIS for the three IQ Scales were Verbal .96, Performance .83 and Full Scale .95 with the WAIS yielding somewhat higher IQs than the WISC-R. Correlations with the WPPSI IQ Scales were Verbal .80, Performance .80 and Full Scale .82. The mean WPPSI IQs were slightly higher than the WISC-R IQs.

Research with Short Forms of the WISC and WISC-R

Most of the research with the WISC short forms has been in the form of validation study whereby the full-length scale was administered to a group of subjects after which certain subtest scores were combined and their total scores correlated to the full scale IQ. The correlation is a measure of the degree to which the short form is a satisfactory estimate of intellectual functioning. All the studies cited have obtained their data from one administration of the full-length scale.

The short form definitely has some limitations and Caldwell and Smith (1968) have suggested that possibly the short forms of the WISC may not be suitable for use with Negro children because of the fact that the intercorrelations of the subtests were found to be lower among Negroes than the standardization group. Hutt and Gibby (1965) stated that the use of the short form should never be considered in any evaluation of a retarded child. On the other hand, the short form was recommended for use with handicapped children (Nickols & Nickols,

1963). With a careful selection of a suitable short form they were able to get reliable results and ensure a positive experience for the child.

Using his abbreviated procedure for the WISC, Yudin (1966) obtained correlations with a group of emotionally disturbed children. Correlations for subtests ranged from .76 to .94 while those for Verbal, Performance and Full Scale scores ranged from .93 to .97. Reid, Moore and Alexander (1968) used Yudin's procedure with a group of brain-damaged and mentally retarded children and they reported the same satisfactory results. In a later study Gayton, Wilson and Bernstein (1970) did not get such unequivocal results using the Yudin procedure. Subjects were children in an outpatient psychiatric clinic. Correlations with the full scales were satisfactory for individual subtests but not for the Verbal-Performance discrepancy factor which they considered an important part of their evaluation.

From the results of a study with 145 emotionally disturbed children, Enburg, Rowley and Stone (1961) highly recommended the use of 30 WISC short forms based on correlations with the full scale IQ of .92 for three-subtest combinations, .94 for four-subtest combinations and .95 for five-subtest combinations. They concluded that the WISC subtest combination short forms were a suitable and reliable estimate of intellectual functioning for children with emotional problems but discouraged their use with problems of organic involvement and/or mental retardation.

In another study with disturbed children, Nickols and Nickols (1963) reported that the WISC short form consisting of Information,

Arithmetic, Digit Span and Picture Completion (or Similarities if there was a visual handicap) had a good correlation (.92) with the full scale when used with disturbed children ranging in age from 6 to 16, and in IQ from 51 to 144. There was a slight tendency for the short form to over-estimate the IQ at the higher levels. Similar results had previously been reported for the short forms of the WAIS when used with an adult schizophrenic population (Nickols, 1962).

Clements (1965) examined 92 children between the ages of 9-0 and 12-11 who had been referred to a clinic for possible reading disabilities. Using a four-subtest combination short form of Similarities, Object Assembly, Arithmetic and Picture Arrangement, correlations of .947 with the full scale were obtained.

In general, the results of the studies using short forms of the WISC and WISC-R suggested that the abbreviated tests provided good estimates of the full scale IQ in many circumstances and with various specialized groups. However, the limitations of the research on the short form made it clear that identifying the kind of decision that has to be made was of primary importance (Levy, 1968).

"Subtest Combination" Short Form

McNemar (1950) set the pattern for studies dealing with the short forms of the WAIS and WISC. He was interested in determining the validities of the short form which consisted of various subtest combinations. He set the example by basing his own research on the original standardization sample of the Wechsler Bellevue Scale and by devising a formula for measuring the degree of correlation between the short form and the full-length test which did not require access

to raw data. The formula which utilized data obtained directly from subtest intercorrelation tables presented in the Wechsler manual (1944) was as follows:

$$r_{xy} = \frac{k + \Sigma\Sigma r_{hj}}{\sqrt{k + 2\Sigma r_{gh}}\ \sqrt{n + 2\Sigma r_{ij}}}$$

Where r_{xy} = correlation between the subtest short form and the full-length scale (called the part-whole correlation or the validity coefficient for the designated short form).

k = number of subtests in the short form.

n = total number of subtests (10 for the WISC-R).

$\Sigma\Sigma r_{hj}$ = the sum of the correlations between each of the k subtests and all the other subtests.

Σr_{gh} = the sum of the intercorrelations of the k subtests.

Σr_{ij} = the sum of the correlations between each of the n subtests.

Note: $\sqrt{n + 2\Sigma r_{ij}}$ becomes a constant in all computations.

The results of McNemar's (1950) research was the systematic ranking of the ten best combinations of two, of three, of four and of five subtests from the Wechsler test. The correlation coefficients of the various short forms with the full-length test tended to be somewhat lower than those which had been published in earlier studies (Hunt, 1948; Patterson, 1948; Rabin, 1943) which was accounted for by the fact that the standardization sample was a much more homogeneous group compared to the samples used in other studies.

"Split-half" Short Form

Another kind of short form to gain acceptance by some researchers (Mogel & Satz, 1963; Pauker, 1963; Satz & Mogel, 1962; Yudin, 1966) was the "split-half" abbreviation of the WAIS and WISC, so called because only half of each subtest was administered. This short form was considered to be a superior form for clinical purposes because it provided a complete breadth of intellectual functioning by sampling each subtest. It was thought to have all the advantages of the full-length test and all the conveniences of a short form. By comparison, the subtest combination short forms were considered to be of limited use because of their emphasis on a single IQ index (Mogel & Satz, 1963).

The split-half procedure was first introduced by Wolfson and Bachelis (1960) who shortened the WAIS Verbal Scale by administering the odd items only on all of the verbal subtests except for the Digit Span subtest which was left unchanged. Correlations between the split-half scores on the Verbal Scale and the full Verbal Scale were reported to be .97.

The split-half form was further developed to include the Performance Scale as well as the Verbal Scale with very encouraging results. The number of items was reduced by as much as 54 percent compared to the full length test and part-whole correlations were reported to be .99 for Verbal IQ, .97 for Performance IQ and .99 for Full Scale IQ (Satz & Mogel, 1962). Following this success a procedure for accurately converting scores on the short form to IQ equivalents was established and the preparation of tables for the

WAIS and WISC (Satz & Mogel, 1962; Yudin, 1966; Silverstein, 1968a) made the use of the split-half short form both sound and practical.

Yudin (1966) proposed a split-half variation of the WISC which consisted of administering every third item on Information, Vocabulary and Picture Completion (scores were later multiplied by 3), every even item on Arithmetic and every odd item on Comprehension, Similarities, Picture Arrangement, Object Assembly and Block Design (scores multiplied by 2), and all items on Digit Span and Coding. Scores were corrected as indicated in the brackets after which the manual could be used to compute scaled scores and IQ's in the usual way. Correlations for this form with the full-length test were Verbal IQ, .96, Performance IQ, .95 and Full Scale IQ, .96. An outline of this short form is presented in Appendix A.

Yudin's (1966) abbreviated procedure for the WISC (and more currently the WISC-R) gained some acceptance particularly for use with specialized groups for whom a short form was more suitable. In a study with brain damaged and mentally retarded children the procedure was found to be very satisfactory (Reid, Moore & Alexander, 1968). In another study (Satz, Van de Riet & Mogel, 1967), the results led the authors to conclude that the Yudin procedure was not suitable for children with above normal IQs.

Although popular in some areas, the split-half procedure shortly began to receive some serious criticism for its statistical claims. A review of the research on the split-half abbreviations (Zytowski & Hudson, 1965) brought attention to the fact that the part-whole correlations in the studies they reviewed were almost all higher

than the odd-even reliability estimates given in the WAIS and WISC manuals. They questioned how a short version of a test could be correlated more highly with the full-length test than the full-length test was correlated with itself. The validity coefficient between part and whole scales could not in reality be higher than the internal consistency coefficient reported by Wechsler. The spuriously high part-whole correlations claimed by the split-half procedure were attributed to a contamination of the results when an abbreviated scale was embedded in the full-length scale. For this reason the use of a profile interpretation obtained from the split-half procedure was strongly discouraged (Zytowski & Hudson, 1965).

Other researchers became highly critical of the split-half short form in the area of test reliability pointing out that short forms which reduced the length of a subtest instead of reducing the number of subtests were ignoring the depreciating effects on reliability (Tellegen & Briggs, 1967). Since reliability is a function of test length, when the number of items in a test is reduced the reliability of the test is decreased. Therefore, it was concluded that the short forms which used fewer items while retaining each subtest (Satz & Mogel, 1962; Yudin, 1966) were ill-suited for the very purpose for which they were designed, namely profile analysis (Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). The same studies also underlined the fact that because the differences between subtest scores in general tend to be much less reliable than the tests themselves it was quite possible that while the total scores might be sufficiently reliable, the profile data would be decidedly unreliable.

It was finally concluded that the split-half abbreviated tests could be used, like the subtest combination tests, as a dependable measure of an IQ index only.

Correlations between the Short Form and the Full-Length Test

The contamination finding which is reflected in the high correlations claimed for both split-half and subtest combination short forms with the criterion test is dealt with by several researchers (Bashaw & Anderson, 1967; Levy, 1967; Silverstein, 1968; Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). The importance of the part-whole correlation cannot be over-rated since it specifies the degree of equivalence between the short form and the full-length test (Tellegen & Briggs, 1967) and is a measure of the validity of the abbreviated test (McNemar, 1950).

When the part-score is obtained from the same test administration as the full-length test score, which is generally the case for most studies (one rare exception is Zytowski and Hudson, 1965), the part-whole correlation cannot be obtained in the traditional manner (Tellegen & Briggs, 1967). It was generally acknowledged that the dependence of the errors in the two sets of data greatly increases the correlation between them but for the wrong reason—incorporated in the correlation coefficient is the perfect correlation between the errors in each test giving it a spuriously inflated value (Levy, 1967). The next logical step in the research process was to define the way in which the part-whole correlations were spuriously high and to devise a formula for determining the validity of a short form which

would be more realistic (Bashaw & Anderson, 1967; Levy, 1967; Tellegen & Briggs, 1967).

As an alternative to the part-whole correlation some researchers suggested and supported the suitability of the part-remainder correlation as a more meaningful validity measure for the short form (Zytowski & Hudson, 1965). The rationale for this consideration was based on the method for split-half correlations. When the part score, that is, the abbreviated form was correlated with the remainder of the test instead of with the whole test, the part-whole error variance was eliminated. The result was a lower and therefore more realistic measure of the part-whole relationship (Zytowski & Hudson, 1965). Other researchers disagreed; while this procedure eliminated error overlap, it also eliminated true-score overlap. For this reason the part-remainder correlation was considered an inappropriate index since it was not by definition a measure of the part-whole relationship (Tellegen & Briggs, 1967).

This evaluation was elaborated further by studies which suggested the probability that the part-whole correlation outlined by McNemar (1950) and the part-remainder correlation represented respectively the upper and lower boundaries of the "true coefficient" (Levy, 1967). Only in cases where the reliability of the part score was very high would the true coefficient approach the part-whole correlation coefficient and only when the reliability was very low would the true coefficient approach the part-remainder correlation coefficient. The true coefficient would lie somewhere between these two coefficients.

Modified Formula for Short-Form Validity

After further studies a formula was developed for obtaining a more accurate measure of the part-whole relationship (Levy, 1967). This index which corrected for the covariance due to correlated error, that is, it excluded part-score error variance, was as follows:

$$r_{xy}' = r_{xy} - (1 - r_{xx})\frac{\sigma_x}{\sigma_y}$$

Where r_{xy}' = corrected part-whole correlation coefficient.

r_{xy} = part-whole correlation coefficient (between short form and full-length test).

r_{xx} = reliability coefficient of each subtest.

$\frac{\sigma_x}{\sigma_y}$ = ratio of error variance to part-score variance.

Note: $(1 - r_{xx})\frac{\sigma_x}{\sigma_y}$ = error variance of the part-score.

This same formula which appeared to solve the problem of spuriously high part-whole correlations was arrived at independently by Bashaw and Anderson (1967) at about the same time.

Tellegen and Briggs (1967) without reference to Bashaw and Anderson (1967) or Levy (1967) approached the problem from a slightly different point of view by suggesting a modified index which replaced the perfect correlation between the part and itself which was assumed in McNemar's (1950) formula, by its reliability coefficient. The result was the following formula:

$$r_{xy}' = \frac{\Sigma r_{hh} + \Sigma\Sigma r_{hj}}{\sqrt{k + 2\Sigma r_{gh}}\ \sqrt{n + 2\Sigma r_{ij}}}$$

Where r_{xy}' = modified coefficient of correlation between composite part and composite whole. (Modified part-whole

correlation.)

Σr_{hh} = sum of reliabilities of the subtests comprising the part.

$\Sigma\Sigma r_{hj}$ = sum of all the intercorrelations between each of the k subtests and all the other subtests.

Σr_{gh} = sum of correlations between any subtests g and h in the part.

Σr_{ij} = sum of correlations between any subtests l and m belonging to the whole.

k = number of subtests included in the part.

n = number of subtests included in the whole.

Note: $n + 2\Sigma r_{ij}$ = becomes a constant in all calculations.

Although it is not immediately apparent to the non-statistician, the relationship of the above formula to that put forward by Bashaw and Anderson (1967) and Levy (1967) is that they are algebraically equivalent (Silverstein, 1971).

In conclusion to their study, Tellegen and Briggs (1967) hypothesized that if the modified index was used to reanalyse the standardization data, the results would yield coefficients lower than those reported by Doppelt (1956) for the WISC subtest short forms and would have the effect of altering the rank order and selection of the best short forms of each combination as determined in several earlier studies using the uncorrected part-whole formula (Clements, 1965; Enburg, Rowley & Stone, 1961; Howard, 1959; Jones, 1962; Maxwell, 1957; McNemar, 1950).

Following this hypothesis suggested by Tellegen and Briggs

(1967) and using the modified formula, the WISC standardization data, obtained from the Wechsler manual, were reanalysed by Silverstein (1970). The results of this study appear in Appendix B. Correlations with the full scale test of the ten best short forms for each combination of two, three, four and five subtests are given in descending order. The values given by the corrected formula averaged .032 lower than those obtained by the old formula (McNemar, 1950). In addition, it was found that 46 of the 120 best short forms selected differed from those selected by the McNemar formula (Silverstein, 1970). This supported the predictions made by Tellegen and Briggs (1967). The new reappraised values were considered to be a more realistic measure of the validity of the subtest short form (Tellegen & Briggs, 1967; Levy, 1967; Bashaw & Anderson, 1967) and the new list of best combinations determined by these validities was deemed to be the most accurate (Silverstein, 1970).

However there was a dissenting note to Silverstein's (1970) summation from McNemar (1974) who maintained that since the whole purpose was to select those short forms which correlated highest with the full scale test, then the original formula (McNemar, 1950) was the most suitable because it accomplished exactly that. Sattler (1974) in the revised edition to his book, appears to have been persuaded by this evaluation to some extent when he chose to revert back to the old uncorrected formula for analysing the standardization data for the new WISC-R. On the other hand, for the results on the WISC and WPPSI, he retained those obtained by Silverstein (1967a, 1967b, 1968, 1970) using the newer corrected formula.

The present writer felt that the argument made for a modified index was reasonable and well documented. The corrected formula should have been used for the new WISC-R data, if only for one important reason, that it makes it possible to compare the best short forms of the WISC and WISC-R. Although some indices obtained may differ only slightly from the old indices, the fact that a better theoretical method has been developed was thought to be a good reason to move on. For this reason, this study has included a reevaluation of the WISC-R data using the corrected formula which provided an updated and more reliable list of validity coefficients for the proposed WISC-R short forms.

In summary, there was much evidence to support the use of a corrected formula for computing part-whole correlations when both scores are obtained in a single administration. The results obtained with the two formulae were considered to be sufficiently different to warrant a reappraisal of the WISC-R standardization data using the corrected formula.

CHAPTER III

RESEARCH DESIGN AND PROCEDURE

The Problem

It was the purpose of this paper to analyse the standardization data of the WISC-R following the method described by Silverstein (1970). Following that analysis and using the resulting list of best Short Forms, a multiple correlational study was carried out, using clinical data obtained from 182 WISC-R protocols, to determine the reliability of the best subtest short forms for use with this particular population.

The Subjects

The subjects were 182 elementary and junior high students, aged 6 1/2-16 1/2, who were enrolled in schools in the Edmonton area and in some smaller rural communities outside the city. All the subjects were tested between September 1975 and July 1976 as part of a University testing program whose purpose was to offer its services to the community while training new psychologists. The protocols were administered by graduate students and examined by clinic supervisors for errors and ommissions. Only those subjects with scores on all ten subtests were included. This study was based on the assumption that the administration and scoring of the protocols was accurate and that the estimates for full scale IQ are reliable and valid.

A characteristic of the group was that all subjects were referred for testing invariably because of some difficulty with their

school work or coping with the school situation. Because all subjects were experiencing a learning problem or coping with a social or emotional problem it was expected that norms for the research group would be lower than the norms for the standardization group.

The Procedure: Part 1

An important preliminary part of this study was to re-analyse the standardization data of the WISC-R using the revised formula (Tellegen & Briggs, 1967) in order to determine the validity coefficients for all possible short forms. It was felt that only then could the short forms of the WISC-R be compared in any meaningful way to those of the WISC which had been obtained statistically by this same method (Silverstein, 1970). The WISC short forms are listed in Appendix B. Also, the results of this analysis yielded the list of forty best short forms which were tested on the research data to obtain a measure of their reliability in the clinic sample.

The formula and method used for the foregoing analysis has been fully elaborated and discussed earlier in this paper. Briefly, the first part of this study was a replication of the analysis of the standardization data of the WISC carried out by Silverstein (1970) using instead the WISC-R data. All necessary data for use with this formula were available in the standardized tables published in the Wechsler Manual (1974). Using this procedure all possible combinations of two, three, four and five subtests were correlated with the full scale IQ and then rank ordered to obtain the ten best short forms in each group. (The total number of possible combinations of the ten subtests is 627; there are 45 different short forms when selecting any

two subtests, 120 when selecting any three subtests, 210 when selecting any four subtests and 252 when selecting any five subtests (McNemar, 1950).)

Two subtests, Digit Span and Mazes were not considered for inclusion in any possible short forms. Both these subtests were excluded in the computation of IQs for the standardization data and for this reason were considered inappropriate for inclusion in a short form which attempts to predict that IQ.

The Procedure: Part 2

The second part of the study was a systematic analysis of the reliability of the 40 short forms to determine the extent to which these short forms were suitable for use with the specialized research group.

Data Analysis

Means, standard deviations and partial correlations were obtained. A step-wise multiple-regression procedure with double cross validation (Mulro 8) was used with the clinic sample to obtain predictive estimates for use with combinations of WISC-R subtests. Subjects were randomly assigned to one of two groups (A and B) for each short form. By this procedure a regression equation was developed for group A and then applied to group B to calculate validity (cross validation) after which the procedure was reversed giving a double cross validation analysis. Groups were randomly re-assigned for each short form analysed.

CHAPTER IV

RESULTS

The Research Sample

A total of 182 protocols of the WISC-R were used in the study. The subjects were 126 male and 56 female students ranging in age from 6 to 16 years. A summary of the percentage of males and females in each age group is presented in Table 1.

TABLE 1

CLINIC SAMPLE BY AGE AND SEX: PERCENT IN EACH AGE GROUP

Age Group	Percent in Each Age Group		
	Male	Female	Total
6-0 to 6-11	7.1	3.8	11.0
7-0 to 7-11	5.5	2.2	7.7
8-0 to 8-11	4.4	2.2	6.6
9-0 to 9-11	7.1	2.2	9.3
10-0 to 10-11	3.8	1.6	5.5
11-0 to 11-11	1.1	1.1	2.2
12-0 to 12-11	2.7	3.8	6.6
13-0 to 13-11	18.1	6.0	24.2
14-0 to 14-11	10.4	3.3	13.7
15-0 to 15-11	7.1	3.3	10.5
16-0 to 16-11	1.6	1.1	2.7
Total	69.2	30.8	100.0

Approximately 70 percent of the sample were boys and 30 percent were girls. This proportion for the clinic group is a reflection of the general finding that many more boys than girls have learning difficulties.

The largest represented age group was 13-0 to 13-11 which accounted for 24 percent of the group. This is the age at which it becomes increasingly difficult for a child to compensate for learning disabilities and/or deficits. Children aged 13-0 to 15-11 accounted for 68 percent of the sample with over two-thirds of those boys. The next highest represented age group was 6-0 to 6-11 accounting for 11 percent of the sample.

A breakdown of the percentage of subjects in each IQ classification compared to the standardized group is summarized in Table 2.

TABLE 2

CLINIC SAMPLE BY IQ CLASSIFICATION COMPARED TO STANDARDIZATION SAMPLE: PERCENT IN EACH CLASSIFICATION

	Standardization Group	Research Group
Mentally Deficient (-69)	2.2	7.1
Borderline (70-79)	6.0	15.9
Dull Normal (80-89)	16.5	24.7
Average (90-109)	49.4	45.0
Bright Normal (110-119)	16.5	5.4
Superior (120-129)	7.4	1.6
Very Superior (130+)	2.3	0.0
Total	100.3	99.7

Because of the specialized nature of the research group IQs tended to be generally lower than those in the standardization group, that is, there were a larger number of subjects with below average IQ and a smaller number with average or above average IQ in the research group. This was an expected result.

The means and standard deviations for the three IQ indices

were all lower for the research group than the standardization sample. These results are shown in Table 3. Compared to the standardized mean of 100 with a SD of 15, the mean Verbal IQ was 87.86 (SD 15.85), Performance IQ was 92.44 (SD 15.77) and Full Scale IQ was 89.07 (SD 15.67).

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR VERBAL IQ, PERFORMANCE IQ AND FULL SCALE IQ FOR THE CLINIC SAMPLE (N = 182)

	Mean	SD
Verbal IQ	87.86	15.85
Performance IQ	92.44	15.77
Full Scale IQ	89.07	15.67

Note: Mean and SD for the Standardized distributions for all three IQ indices are 100.0 and 15.0.

Scaled scores for every individual subtest were lower than the standardized mean of 10 while standard deviations were usually somewhat higher than the standardized 3. Similarly Sums of Scaled Scores for Verbal, Performance and Full Scale were lower, 40.38 (SD 12.92) compared to 50.25 (SD 12.14), 44.64 (SD 11.80) compared to 50.19 (SD 10.89) and 85.02 (SD 22.76) compared to 100.44 (SD 21.01). These results are summarized in Table 4.

Results for Part 1: 40 Best Short Forms

The standardization data for the WISC-R were analyzed according to the method suggested by Silverstein (1970) to determine

TABLE 4

WISC-R MEANS AND STANDARD DEVIATIONS OF SCALED SCORES FOR THE CLINIC SAMPLE (N = 182)

	Mean		SD	
Information	7.75	(10.00)*	3.03	(3.00)*
Similarities	8.26		3.18	
Arithmetic	7.87		3.01	
Vocabulary	8.08		3.10	
Comprehension	8.42		3.00	
Sum of Scaled Scores— Verbal	40.38	(50.25)	12.92	(12.14)
Picture Completion	9.31	(10.00)	3.05	(3.00)
Picture Arrangement	9.36		3.58	
Block Design	8.50		3.20	
Object Assembly	9.56		2.86	
Coding	7.88		3.17	
Sum of Scaled Scores— Performance	44.64	(50.19)	11.80	(10.89)
Sum of Scaled Scores— Full Scale	85.02	(100.44)	22.76	(21.01)

Note: *Figures in parentheses are the mean scaled scores of the subtests and sum of verbal, performance and full scale scores of the standardization sample.

which short forms correlated best with the full scale. The 10 best short forms in each combination of two, three, four and five subtests were determined and are listed with their validity coefficient in Table 5.

The 40 best short forms which were recommended by Silverstein (1970) for the WISC are marked with an asterisk; there are 22 short forms which appear on both lists. An informative comparison can now be made with Table 5 by referring to the complete list of 40 best short forms for the WISC in Appendix B.

More than half the WISC-R short forms are the same as those selected for the WISC which is an expected finding considering the high degree of similarity between the two tests and the continued high subtest reliabilities of the subtests Vocabulary, Information, Similarities and Block Design which appear so frequently in the short forms. The degree to which the two sets vary can be explained by the improved reliability of some of the subtests of the WISC-R, in particular, Comprehension and Picture Completion. Comprehension, for example, had a reliability range of .59 to .73 on the WISC compared to .69 to .87 on the WISC-R; Picture Completion had a reliability range from .59 to .69 on the WISC compared to .68 to .85 on the WISC-R. As a result these two subtests appeared with more frequency in the WISC-R short forms: Comprehension appeared 10 times compared to 5 on the WISC; Picture Completion appeared 11 times compared to twice on the WISC. These subtests replaced in frequency other subtests which appeared in fewer of the WISC-R short forms: Arithmetic appeared in only 4 compared to 15 short forms of the WISC; Picture Arrangement,

TABLE 5

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS:
A REPLICATION OF THE PROCEDURE RECOMMENDED
BY SILVERSTEIN (1970)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
*V BD	.882	*I V BD	.903	I V PC BD	.916	I S V PC BD	.926
*I BD	.862	*S V BD	.902	*I S V BD	.914	I V C PC BD	.925
S BD	.856	*I S BD	.896	*S V PA BD	.912	*I S V PA BD	.925
C BD	.846	*I C BD	.896	I V C BD	.911	*S A V PA BD	.924
*V OA	.839	*V C BD	.894	I S C BD	.911	*I S V OA BD	.923
*S V	.838	V PC BD	.887	*I V PA BD	.911	I S C PC BD	.922
V PC	.836	*S C BD	.887	S V PC BD	.910	I V PC PA BD	.922
*I V	.834	*A V BD	.886	*S A V BD	.909	I S C PA BD	.922
I S	.832	V PA BD	.885	I C PC BD	.909	I V PC CO BD	.922
I PC	.825	*S V OA	.881	I V OA BD	.908	*S A V OA BD	.922

* Short Forms which also appear on the WISC list of best 40 (Silverstein, 1970).

Note 1: Abbreviations: I = Information; C = Comprehension; A = Arithmetic; S = Similarities; V = Vocabulary; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: SE_{est} for Dyad range from 7.08-7.59; Triad 6.44-7.09; Tetrad 6.01-6.29; Pentad 5.66-5.81.

in 7 compared to 14; Object Assembly, in 5 compared to 9 short forms of the WISC. The Coding subtest appeared in only one short form for each test.

The subtests appearing most frequently in the short forms of the WISC-R were those with the highest reliability: Block Design (.85), Vocabulary (.86), Information (.85) and Similarities (.81); respectively they were represented in 33, 28, 22 and 19 of the 40 short forms. All except Vocabulary appeared more frequently (3 to 5 times) in the WISC-R tests.

A further comparison was made between the results in Table 5 and those obtained by the McNemar (1950) method, which was a comparison of the results obtained by two different procedures. The latter were reported in Sattler (1974) and are presented in Appendix C. While varying noticeably in validity coefficient values, the two lists were very similar in selection of short forms. Of the 40 short forms, 29 are the same for both lists; of those which changed, 6 were pentads, 4 were tetrads and 1 was a triad; all dyads remained unchanged.

The two lists of short forms are similar to the degree that the results are largely dependent on the reliabilities of the individual subtests. Both methods used the reliability coefficients reported in the manual (Wechsler, 1974). The extent of their variability can be accounted for by the different procedures used. In general, the corrected formula yielded lower validity coefficients for the short forms, an outcome which was expected and preferred as being a more realistic value (Tellegen & Briggs, 1967; Silverstein, 1970).

The standard errors of estimate (SE_{est}) for the WISC-R were

calculated as 7.08 IQ points for the best dyad, 6.44 for the best triad, 6.01 for the best tetrad and 5.66 for the best pentad. It should be noted that these values are somewhat higher than those reported by the other method (refer to Appendix C). Because SE_{est} is a function of the validity coefficients their value increased when the coefficients were decreased. As a result, in addition to a more realistic validity coefficient, the correct formula yields what is probably a more realistic standard error of measurement.

The list of correlation coefficients presented in Table 5 is intended to replace the list shown in Sattler (1974) which was considered inappropriate for use in comparing the WISC and WISC-R. By referring to Table 5 and Appendix B which lists the 40 best WISC short forms, it was observed that indeed the WISC-R short forms, largely because of improved subtest reliability as mentioned earlier, were a more valid measure of IQ than those of the WISC. The validity coefficients in general, tended to be moderately, but consistently, higher compared to the WISC. For example, for the best dyad (V BD) the validity coefficient was .882 compared to .856, for the best triad (I V BD) it was .903 compared to .887, for the best tetrad (I V PC BD) it was .916 compared to .904 and for the best pentad (I S V PC BD) it was .926 compared to .915. Validities were improved to the extent that for each group the new range of validity coefficients was equivalent to the range of the next largest group on the WISC, for example, the validity for the dyads was equal to that of the triads on the old test. This finding supports the general conclusions concerning the WISC-R, that it is an even more rigorous test than its predecessor

(Sattler, 1974; Wechsler, 1974).

Results for Part 2: Cross Validation Analysis

The data from 182 protocols were used to evaluate the reliability of the 40 best WISC-R Short Forms. This analysis was obtained by a multiple correlation procedure with a double cross validation procedure. For each short form analysis subjects were randomly reassigned to one of two groups and regression equations calculated for each group. The results were expressed by the squared multiple correlation coefficient (R^2) which gave a measure of the proportion of the variability in the Full Scale IQ which could be accounted for by the regression equation of each particular short form. The results of this analysis are summarized in Table 6.

Three R^2 calculations were made for each short form: The first (R^2) was a measure of the relationship between the short form and the full-length test for the research sample; the second (R'^2) was an unbiased estimate of R^2 for the population calculated according to Olkin and Pratt (1958) and Tatsuoka (1969). The third was the cross validated R^2 whereby scores calculated on Group 1 from the regression equations for Group 2, and on Group 2 from regression equations for Group 1, were correlated with the Full Scale IQ. This final statistic was the strongest one that could be made when both predictor and criterion variables were obtained at one time. The results in general were very strong and appear to overwhelmingly support the efficacy of many of the WISC-R short forms. Theoretically, R^2 would be greater than R'^2 or the cross-validated R^2; R'^2 would be smaller than R^2 because it was an estimate of the population; cross-validated R^2 would

TABLE 6

CROSS VALIDATED MULTIPLE CORRELATION COEFFICIENTS (R^2) FOR 40 WISC-R SHORT FORMS OBTAINED FROM RESEARCH SAMPLE
(N = 182)

	Group 1			Group 2		
Short Form	R^2	R'^2	R^{2*}	R^2	R'^2	R^{2*}
V BD	.85	.85	.85	.80	.80	.80
I BD	.85	.85	.84	.82	.82	.81
S BD	.75	.75	.73	.75	.75	.71
C BD	.80	.80	.78	.74	.74	.72
V OA	.81	.81	.81	.81	.81	.80
S V	.75	.75	.75	.75	.75	.75
V PC	.80	.80	.80	.81	.81	.81
I V	.78	.78	.78	.78	.78	.78
I S	.79	.79	.79	.76	.76	.76
I PC	.82	.82	.82	.77	.77	.77
I V BD	.89	.85	.88	.88	.83	.87
S V BD	.86	.83	.85	.86	.83	.85
I S BD	.89	.86	.89	.84	.80	.84
I C BD	.88	.82	.88	.87	.84	.87
V C BD	.84	.80	.83	.86	.86	.86
V PC BD	.90	.86	.89	.82	.78	.81
S C BD	.87	.80	.87	.76	.67	.75
A V BD	.91	.86	.91	.82	.78	.82
V PA BD	.82	.76	.82	.93	.89	.92
S V OA	.76	.71	.76	.90	.87	.90
I V PC BD	.90	.90	.89	.92	.92	.91
I S V BD	.90	.90	.87	.90	.90	.88
S V PA BD	.88	.88	.88	.93	.93	.93
I V C BD	.89	.89	.89	.89	.89	.89
I S C BD	.89	.89	.88	.89	.89	.89
I V PA BD	.92	.92	.91	.92	.92	.89
S V PC BD	.91	.91	.90	.88	.88	.87
S A V BD	.92	.92	.91	.88	.88	.86
I C PC BD	.93	.93	.92	.89	.89	.88
I V BD OA	.92	.92	.91	.92	.92	.91
I S V PC BD	.93	.90	.92	.92	.92	.92
I V C PC BD	.92	.90	.92	.93	.92	.93
I S V PA BD	.93	.92	.92	.94	.93	.93
S A V PA BD	.94	.92	.93	.95	.93	.94
I S V BD OA	.92	.91	.90	.94	.93	.93
I S C PC BD	.94	.91	.94	.92	.90	.91
I V PC PA BD	.93	.92	.92	.94	.92	.93
I S C PA BD	.93	.91	.93	.93	.92	.93
I V PC BD CO	.93	.91	.93	.94	.92	.94
S A V BD OA	.92	.91	.92	.93	.91	.93

Note:

R^2 = the squared multiple correlation coefficient of the subtest with the Full Scale IQ.

R'^2 = an unbiased estimate of the squared multiple correlation coefficient for the population. It is calculated according to Olkin and Pratt (1958) and Tatsuoka (1969) by a formula which corrects for shrinkage.

R^{2*} = the cross validated R^2, calculated for Group 1 from the regression equation constructed for Group 2, and for Group 2 from the regression equation constructed for Group 1. It is a measure of predictive efficiency.

be expected to be lower than either R^2 or R'^2. In fact, cross validated R^2 was equal to and often exceeded the other coefficients which is to say the actual data yielded higher coefficients than would be theoretically expected.

It should be mentioned again that the values presented in Table 6 are the squared multiple correlations and express the variance which can be accounted for by each short form in predicting full scale IQ. With this consideration in mind some noteworthy observations can be made. Dyads have a wide variability; at their worst they accounted for only 71 to 73 percent of the variance in predicting IQ, at their best they accounted for 81 to 85 percent. The triads showed much the same kind of variability as the dyads but at their best they could make a prediction which accounted for 91 to 92 percent of the variance. When the short forms were increased from three to four subtests, the variability was reduced considerably. The tetrads were able to predict full scale IQ which accounted for 86 to 93 percent consistently. The five-subtest short form compared favorably with the full scale test by accounting for 90 to 94 percent of the variance when predicting IQ.

In general, the reliability of the short forms increased as the length of the short form increased. It was observed that the range of correlation coefficients narrowed as the short form lengthened indicating less variability and better reliability for the longer test. For example, when the results for Group 2 were examined, it was found that multiple R^2 correlations for the two-subtest short forms ranged from .71 to .81, for the three-subtest short forms from .75 to .92, for the four-subtest short forms from .86 to .93 and for the five-

subtest short forms from .91 to .94.

The ability of the short form to predict the full scale IQ successfully was increased as the number of subtests in the short form was increased. However, while the differences in reliability between the dyad and triad, and between the triad and tetrad were quite noticeable, the differences between the tetrad and pentad were not as great. For example, the best multiple R^2 correlations for a dyad were .85 and .81 (highest for each group) compared to .91 and .92 for a triad, .92 and .93 for a tetrad and .94 and .94 for a pentad. It was concluded that the four-subtest short forms were the most elegant short forms giving a good reliable estimate of IQ for the very reasonable amount of time required.

The evaluation of the efficacy of the short forms in predicting Verbal IQ and Performance IQ has been summarized in Table 7.

In general, the short forms were better predictors of Verbal IQ (V IQ) than Performance IQ (P IQ). For example, the multiple R^2 correlation coefficients for the dyads ranged from .70 to .91 for V IQ compared to .43 to .70 for P IQ; for the triads, from .82 to .91 for V IQ compared to .58 to .81 for P IQ; for the tetrads, .88 to .96 for V IQ compared to .68 to .85 for P IQ and in the pentads, .90 to .95 for V IQ compared to .73 to .91 for P IQ. The variability for each group of short forms was greater when predicting P IQ. For example, in the most reliable group, the pentads, the multiple R^2 ranged from .73 to .91 for P IQ compared to .90 to .95 for V IQ.

The short forms also were able to predict Verbal IQ somewhat better than Full Scale IQ (FS IQ) (see Table 6; compare Table 5). For

TABLE 7

CROSS VALIDATED MULTIPLE CORRELATION COEFFICIENT (R^2) FOR VERBAL IQ AND PERFORMANCE IQ FOR CLINIC SAMPLE (N = 182)

Short Form	Verbal IQ		Performance IQ	
	Group 2 per Group 1 Equation	Group 1 per Group 2 Equation	Group 2 per Group 1 Equation	Group 1 per Group 2 Equation
1	.81	.85	.66	.73
2	.77	.80	.69	.73
3	.70	.72	.69	.64
4	.79	.74	.67	.66
5	.87	.78	.55	.70
6	.87	.92	.32	.45
7	.79	.84	.66	.70
8	.87	.91	.43	.43
9	.90	.85	.42	.43
10	.84	.70	.69	.70
11	.88	.92	.71	.65
12	.88	.91	.71	.63
13	.89	.86	.63	.79
14	.90	.91	.63	.79
15	.89	.89	.75	.60
16	.85	.83	.87	.72
17	.86	.88	.76	.58
18	.90	.89	.71	.68
19	.86	.82	.86	.81
20	.89	.91	.68	.61
21	.91	.89	.80	.82
22	.93	.96	.73	.69
23	.88	.91	.81	.85
24	.94	.95	.68	.76
25	.96	.94	.63	.76
26	.90	.90	.81	.82
27	.91	.88	.81	.79
28	.95	.95	.74	.68
29	.91	.88	.85	.78
30	.90	.90	.78	.85
31	.94	.94	.83	.77
32	.92	.94	.82	.79
33	.94	.95	.82	.86
34	.96	.93	.82	.86
35	.95	.94	.82	.82
36	.95	.94	.80	.81
37	.89	.91	.89	.91
38	.93	.95	.79	.86
39	.90	.90	.88	.88
40	.96	.94	.83	.73

example, the range of correlations for the dyads was .71 to .81 for the FS IQ compared to .70 to .91 for V IQ; for triads, .75 to .92 for FS IQ compared to .82 to .91 for V IQ; for tetrads, .86 to .93 for FS IQ compared to .88 to .96 for V IQ; for pentads, .91 to .94 for FS IQ compared to .90 to .95 for V IQ.

For this study multiple regression weights and constants were determined for each short form. These are given in Table 8. Since the results of Group 1 and 2 were so similar throughout the analysis, it was decided that only the data from Group 1 would be presented. The regression equation for computing the estimated criterion variable (Y') from the predictor variables (b_1, b_2, . . . b_5) is shown at the bottom of the table.

The multiple regression weights and constants shown in the table are those derived from the sample data and used in the double cross validation analysis. Theoretically the derived regression equations could be used to predict full scale IQs in another sample, particularly in a sample which was similar to the research group. The limitations of the predictions made on this basis are the same as those for any regression equation and that is that the weights and constants are derived from a specific sample and not from the population. However, many of the short forms as determined by this study, have a sufficiently high correlation (see Table 6) to warrant a high degree of confidence in their ability to predict the full scale IQ.

TABLE 8

MULTIPLE REGRESSION EQUATION WITH ADDITIVE CONSTANT (a) AND REGRESSION WEIGHTS FOR USE IN PREDICTING FULL SCALE IQ

	a	b_1	b_2	b_3	b_4	b_5
V BD	44.76	3.08	2.22			
I BD	44.75	2.97	2.49			
S BD	48.26	2.07	2.78			
C BD	42.58	2.44	3.02			
V OA	39.04	3.77	1.98			
S V	47.57	1.95	3.10			
V PC	42.23	3.48	2.05			
I V	49.54	2.68	2.25			
I S	48.24	3.16	1.89			
I PC	41.15	3.56	2.14			
I V BD	42.89	2.14	1.60	1.93		
S V BD	42.89	1.29	2.25	2.02		
I S BD	43.13	2.49	1.47	1.72		
I C BD	41.89	2.43	1.48	1.91		
V C BD	42.20	2.26	1.36	1.99		
V PC BD	39.50	3.01	1.37	1.45		
S C BD	38.18	1.47	2.35	2.28		
A V BD	41.53	1.73	2.50	1.57		
V PA BD	42.36	2.53	1.21	1.80		
S V OA	41.16	1.25	2.55	1.76		
I V PC BD	42.60	2.16	1.29	.95	1.26	
I S V BD	42.63	2.25	.92	.89	1.69	
S V PA BD	39.51	1.10	2.13	1.23	1.41	
I V C BD	43.29	1.88	1.28	.84	1.68	
I S C BD	42.67	2.26	.93	.96	1.61	
I V PA BD	39.51	1.55	1.44	.96	1.98	
S V PC BD	37.72	1.12	1.96	1.30	1.60	
S A V BD	39.85	1.18	1.54	1.62	1.60	
I C PC BD	34.46	2.14	1.37	1.20	1.79	
I V BD OA	37.85	1.87	1.56	1.45	1.24	
I S V PC BD	38.72	1.25	.99	1.53	1.08	1.25
I V C PC BD	37.36	1.34	1.15	1.16	1.28	1.30
I S V PA BD	41.56	1.05	.82	1.36	1.23	1.17
S A V PA BD	40.24	.90	1.09	1.57	1.30	.95
I S V BD OA	35.90	1.55	.72	1.49	1.14	1.40
I S C PC BD	39.40	1.64	.85	1.11	1.03	1.35
I V PC PA BD	41.37	1.54	1.56	.74	.55	1.39
I S C PA BD	37.88	1.64	.95	1.11	1.06	1.42
I V PC BD CO	37.01	1.53	1.51	1.04	1.13	1.08
S A V BD OA	37.65	.86	1.01	1.82	1.10	1.24

Note: $Y' = a + b_1X_1 + b_2X_2 \ldots b_5X_5 + e_i$

CHAPTER V

SUMMARY, DISCUSSION, IMPLICATIONS AND CONCLUSIONS

The results reported in Chapter IV give encouraging support to the thesis of this study which was that the WISC-R short forms are a sound and sensible alternative to the full-length test in many situations and can be considered by the serious and busy test administrator as part of a battery or as a single quick measure of intelligence.

The study explored the spuriously high component in correlations involving two sets of scores obtained in just one administration of a test and the process that was necessary to resolve the problem satisfactorily. The new corrected formula for obtaining part-whole correlations was suggested by Tellegen and Briggs (1967) and considered by Silverstein (1970) to be a sound solution.

The replication of Silverstein's (1970) procedure on the WISC-R standardization data produced a list of 40 best short forms (Table 5) which could be compared and evaluated in a meaningful way to the list of 40 best short forms for the WISC (Appendix B). A comparison of the two lists indicated important similarities and differences.

It was observed that 22 of the short forms appeared on both lists, a finding which was expected since the two full-length tests were so similar. The changes in the remaining 18 short forms of the WISC-R were accounted for by the increased appearance of two subtests,

Comprehension and Picture Completion which, due to improved subtest reliability, tended to preempt the subtests Arithmetic, Picture Arrangement and Object Assembly which appeared more frequently on the old test.

The validity coefficients of the WISC-R short forms were found to be consistently higher than those for the WISC leading to the conclusion that the revised test was more robust than the WISC as suggested by Wechsler (1974), Sattler (1974) and Kaufman (1975).

The results of the cross validation analysis shown in Tables 6-8 indicated that the short forms, in general, were a very reliable prediction of the full scale IQ and could be used with confidence to estimate the intellectual level of a child.

It was noted that the short forms were better predictors of Verbal IQ than Performance IQ, in fact, they were better predictors of Verbal IQ than Full Scale IQ. This finding was explained by noting that verbal subtests, or those correlating highly with the verbal scale, were heavily represented in the short forms. Vocabulary, Information and Similarities were each represented in more than half the short forms while Block Design, which correlates as well with the verbal scale as it does with the performance scale, was represented in 33 of the 40 short forms. The high reliability of these subtests, along with their frequency in the short forms, ensures that the short forms will be excellent measures of Verbal IQ.

The squared multiple correlations for all short forms were higher than theoretically expected. It was found that the two-subtest short form could make a prediction of IQ which accounted for 71 to 85

percent of the variance. As the length of the short form was increased the ability to predict the full scale IQ was improved, so that any three-subtest short form could predict an IQ within a 75 to 92 percent range of variance, a four-subtest short form within an 86 to 93 percent range of variance and any five-subtest short form within the 90 to 94 percent range of variance. These results gave strong support for their use in groups such as the research sample where children were experiencing some school difficulties.

The tetrads or four-subtest short forms were considered to be the most efficient and most useful because of their high predictive ability in relation to the amount of administration time required.

From the list of 40 best short forms of the WISC-R an examiner could select those which suited his needs and preferences. It was suggested that some care in selecting an appropriate short form would ensure the probability of both a good reliable result and a positive experience for the child.

Short forms were not recommended for use with retarded children or where classification decisions were to be made, such as opportunity class placement. Short forms however, should be very useful with children who have been referred to a clinic or counsellor for emotional disturbance or behavioral problems. They were also recommended for use with handicapped children who might find a longer test tiring, and for those who had a possible reading disability (Clements, 1965).

Because some children find it difficult to keep still for very long or for some reason cannot spend a long time on one task, any short form would likely be a better choice over the full-length test.

Individual subtest characteristics and functions were important considerations in selecting a short form. Vocabulary, for example, is a long and demanding task for which many children lack the skill, in which case a short form with Information or Similarities which are shorter tests and seem easier, would be the better choice. Block Design was included in 33 of the 40 short forms and luckily is a test which most children enjoy; errors in this test are often not apparent to the child, so there is no discouragement. For a blind child the only short forms would be those without any of the Performance tests which leaves only three dyads to choose from, Similarities—Vocabulary, Information—Vocabulary and Information—Similarities. For a child who is deaf but able to read, there would be no problem in choosing short forms; if the child could not read, the short forms would not be appropriate. With a behavior problem child, that is, one who was restless or inattentive or not easily engaged, a full-length test would be difficult; a short form which is quick and interesting would be more successful, for example, Information—Comprehension—Picture Completion—Block Design.

Limitations of the Study

The major limitation of this study was that again the issue of the reliability of short forms was decided on data obtained from a single administration of the full-length test. Although the statistics used were the most rigorous available for making corrections under these conditions, the question must be explored further in a study which would use short form data which had been obtained from a separate administration.

Future research could address itself to this problem by testing a group with a short form and some time later with the full-length test. Since both tests would not have to be administered by the same person, the study could be made in cooperation with the Individual Testing psychology course at the University. Short forms could be administered to those who would be or had been tested by students in the program.

Implications of the Study

The purpose of this study was to, firstly, analyse the standardization data of the WISC-R in order to determine the 40 best short forms, and secondly, to determine the reliabilities of these short forms with a clinic sample. The results of the study were positive and unequivocal. In terms of this study the implications are as follows:

1. The list of 40 best short forms of the WISC-R which was determined by this study (Table 5) is intended to be used by all those who do testing to select appropriate short forms according to needs and preferences. In addition, the list is intended to replace a similar list published in Sattler (1974).
2. This study implies that short forms of the WISC-R can be used with confidence as reliable and valid measures of intellectual level in a variety of situations. The short forms work very well for the purpose for which they were intended, that is, a good estimate of intellectual functioning.

3. This study adds to the body of research which supports the use of the WISC-R short forms in general practice.
4. Similarly, it contributes to the research which supports the reputation of the WISC-R as an even more robust test than its predecessor.

Conclusion

The list of suggested short forms derived from this study and tested on the research group appear to have excellent utility for such a specialized group. Whereas previous studies in general supported the use of the short form, there was some question as to whether or not they would be suitable for use with children referred to a psycho-educational clinic.

An evaluation of the data indicated that some short forms had better utility than others. If the criterion was only to save time then any of the two-subtest short forms was clearly the best; if high reliability was the criterion then the five-subtest short forms were the best. However, the three-subtest and four-subtest short forms while only slightly more time consuming were appreciably more reliable than the two-subtest short form and for all practical purposes equally as reliable as the five-subtest short form. For this reason they were considered the most preferred and strongly recommended. Adding a fifth subtest did not increase reliability sufficiently to warrant the required extra administration time, although it could be easily included when circumstances demanded it.

Within the set of triads and tetrads other criteria could certainly be used by the administrator in selecting an appropriate

test. The fact that certain subtests are longer or less interesting for the child while others are quicker and more fun, and the fact that some require special test materials while others are entirely verbal would be determining factors. It was felt that a short form could be selected on the basis of the individual, the environment and the tester according to suitability and preference.

When using the WISC-R Short Forms some caution must be exercised in the interpretation of results. The standard error for the short form is definitely higher than the full-length test. The standard error of the estimate (SE_{est}) for the WISC-R short forms varies from 7.08 IQ points for the best dyad to 5.66 for the best pentads (refer to note in Table 5). The average standard error of measurement (SE_m) for the full scale IQ is 3.19, for Verbal IQ 3.60. Therefore any individual IQ can be over- or under-estimated to a large degree. No single administration of a short form can be interpreted as anything more than a rough estimate. However, this fact was not considered to be a great disadvantage in as much as it was thought that it should reduce the tendency of those interested in labeling a child with a number and thinking of that number as invariable. The explicit assumption of a short form result is that it is an imperfect estimate of general intelligence indicating roughly a child's intellectual development as compared to other children his age.

The WISC-R short forms yield a quick and very reliable estimate of an individual's intellectual level. They will also detect any gross impairments which would indicate the necessity for administering the full-length test or to continue with further

testing. They are a very economical way of obtaining a highly valid measure of intelligence.

There is much evidence to support the use of the short forms with any public school population. However, they are not recommended for diagnosing learning disabilities or for making educational decisions about mental retardates or special education candidates.

The advantages of the short form are that they are practical, save time, serve as good screening devices, allow for more frequent testing and can be administered by less than fully qualified psychologists. A classroom teacher, for example, can use a WISC-R Short Form to test or support a hypothesis concerning a child, almost immediately without too much formality and have the information very soon afterwards.

The various WISC-R short forms are to some degree less valid than the full-length form but not so much so that they cannot be seriously considered for use in many testing situations. Every user of the short form then must consider the nature of the decision to be made and evaluate the relationship between validity lost and time saved.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bashaw, W. L. & Anderson, H. E. Jr. A correction for replicated error in correlation coefficients. *Psychometrika*, 1967, *32*, 435-441.

Blatt, S. J. & Allison, J. The intelligence test in personality assessment. In A. I. Rabin (Ed.), *Projective techniques in personality assessment*. New York: Springer, 1968, 421-460.

Caldwell, M. B. & Smith, T. A. Intellectual structure of southern Negro children. *Psychological Reports*, 1968, *23*, 63-71.

Clements, C. R. An abbreviated form of the Wechsler Intelligence Scale for Children. *Journal of Consulting Psychology*, 1965, *29*, 92.

Doppelt, J. E. Estimating the full scale score on the WAIS from scores on four subtests. *Journal of Consulting Psychology*, 1956 *20*, 63-66.

Draper, M. R. & Smith, H. *Applied regression analysis*. New York: John Wiley & Sons Inc., 1966.

Enburg, R., Rawley, V. N. & Stone, B. Short forms of the WISC for use with emotionally disturbed children. *Journal of Clinical Psychology*, 1961, *17*, 144-149.

Finch, A. J., Ollendick, T. H. & Ginn, F. W. WISC short forms with mentally retarded children. *American Journal of Mental Deficiency*, 1973, *78*(2), 144-149.

Fromm, E., Harman, L. D. & Marschak, M. Children's intelligence tests as a measure of dynamic personality functioning. *American Journal of Orthopsychiatry*, 1957, *27*, 134-144.

Frost, B. P. & Frost, R. The pattern of WISC scores in a group of juvenile psychopaths. *Journal of Clinical Psychology*, 1962, *18*, 354-355.

Gayton, W. F., Wilson, W. T. & Bernstein, S. An evaluation of an abbreviated form of the WISC. *Journal of Clinical Psychology*, 1970, *26*, 466-468.

Glasser, A. J. & Zimmerman, I. L. *Clinical interpretation of the Wechsler Intelligence Scale for Children*. New York: Grune & Stratton, 1967.

Hunt, W. A. The validity of some abbreviated intelligence scales. *Journal of Consulting Psychology*, 1948, *12*, 48-52.

Hutson, B. A. Psychological testing; misdiagnosis and half-diagnosis. *Psychology in the Schools*, 1974, *11*, 388-391.

Hutt, M. L. & Gibby, R. G. *The mentally retarded child* (2nd ed.). Boston: Allyn and Bacon, 1965.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1-123.

Joesting, J. & Joesting, R. Quick test validation scores of adults in a welfare setting. *Psychological Reports*, 1972, *30*, 537-538.

Kaufman, A. S. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. *Journal of Consulting and Clinical Psychology*, 1975, *43*(2), 135-147.

Kelly, F. J., Beggs, D. & McNeil, K. A. *Research design in the behavioral sciences: Multiple regression approach*. Southern Illinois University Press, 1969.

Levy, P. The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, 1967, *23*, 84-86.

Levy, P. Short-form tests: A methodological review. *Psychological Bulletin*, 1968, *69*, 410-416.

McNemar, Q. On abbreviated Wechsler-Bellevue Scales. *Journal of Consulting Psychology*, 1950, *14*, 79-81.

McNemar, Q. *Psychological statistics* (3d ed.). New York: Wiley, 1962.

McNemar, Q. Correction to a correction. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 145-146.

Maxwell, E. Validities of abbreviated WAIS scales. *Journal of Consulting Psychology*, 1957, *21*, 121-126.

Mogel, S. & Satz, P. Abbreviation of the WAIS for clinical use: An attempt at validation. *Journal of Clinical Psychology*, 1948, *12*, 147-152.

Nickols, J. Brief forms of the Wechsler Intelligence Scales for research. *Journal of Clinical Psychology*, 1962, *18*, 167.

Nickols, J. & Nickols, M. Brief forms of the WISC for research. *Journal of Clinical Psychology*, 1963, *19*, 425.

Olkin, I. & Pratt, J. W. Unbiased estimate of certain correlation coefficients. *Annals of Mathematics Statistics*, 1958, *29*, 201-211.

Patterson, C. H. A further study of two short forms of the Wechsler-Bellevue Scale. *Journal of Consulting Psychology*, 1948, *12*, 147-152.

Patterson, C. H. *The Wechsler-Bellvue scales: A guide for counsellors.* Springfield, Illinois: Charles C. Thomas, 1953.

Pauker, J. D. A split-half abbreviation of the WAIS. *Journal of Clinical Psychology*, 1963, *19*, 98-100.

Rapaport, D., Gill, M. M. & Schafer, R. *Diagnostic psychological testing* (Rev. ed.). New York: International Universities Press, 1968.

Reid, W. B., Moore, D. & Alexander, D. Abbreviated form of the WISC for use with brain-damaged and mentally retarded children. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 236.

Rhodes, F. *Manual for the Rhodes WISC scatter profile.* San Diego, California: Educational and Industrial Testing Service, 1969.

Sattler, J. M. *Assessment of children's intelligence* (Rev. ed.). Philadephia: W. B. Saunders, 1974.

Satz, P. & Mogel, S. An abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology*, 1962, *18*, 77-79.

Satz, P., Van de Riet, H. & Mogel, S. An abbreviation of the WISC for clinical use. *Journal of Consulting Psychology*, 1967, *31*, 108.

Schofield, W. Critique of scatter and profile analysis of psychometric data. *Journal of Clinical Psychology*, 1952, *8*, 16-22.

Silverstein, A. B. "Validity" of short-form intelligence tests. *Journal of Consulting Psychology*, 1965, *29*, 392-393.

Silverstein, A. B. A short form of the WISC and WAIS for screening purposes. *Psychological Reports*, 1967, *20*, 682. (a)

Silverstein, A. B. A short form of the Wechsler scales for screening purposes. *Psychological Reports*, 1967, *21*, 842. (b)

Silverstein, A. B. Validity of a new approach to the design of WAIS, WISC, WPPSI short forms. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 478-479.

Silverstein, A. B. Reappraisal of the validity of the WAIS, WISC and WPPSI short forms. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 12-14. (a)

Silverstein, A. B. Reappraisal of the validity of a short form of Wechsler's scales. *Psychological Reports*, 1970, *26*, 559-561. (b)

Silverstein, A. B. A corrected formula for assessing the validity of WAIS, WISC and WPPSI short forms. *Journal of Clinical Psychology*, 1971, *27*, 212-213.

Silverstein, A. B. A short short form of the WISC-R for screening purposes. *Psychological Reports*, 1974, *35*, 817-818.

Silverstein, A. B. Comment: A reply of McNemar. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 423-425.

Tatsuoka, M. M. *Selected Studies in Advanced Statistics. Number 5.* Champaign, Illinois: The Institute for Personality and Ability Testing, 1969.

Tellegen, A. & Briggs, P. F. Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*, 1967, *31*, 499-506.

Tipton, R. M. & Stroud, L. H. Abbreviated forms of the WAIS. *American Journal of Mental Deficiency*, 1973, *78*(2), 150-152.

Wechsler, D. *The Measurement of Adult Intelligence* (3d ed.). Baltimore: Williams & Wilkins, 1944.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children.* New York: Psychological Corporation, 1949.

Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

Wechsler, D. *The Measurement of Adult Intelligence* (4th ed.). Baltimore: Williams & Wilkins, 1958.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children* (rev. ed.). New York: Psychological Corporation, 1974.

Yudin, L. W. An abbreviated form of the WISC for use with emotionally disturbed children. *Journal of Consulting Psychology*, 1966, *30*, 272-275.

Zytowski, D. G. & Hudson, J. The validity of split-half abbreviations of the WAIS. *Journal of Clinical Psychology*, 1965, *21*, 292-294.

APPENDICES

APPENDIX A

ABBREVIATED PROCEDURES OF THE WECHSLER SCALES USING THE "SPLIT-HALF" METHOD

APPENDIX A

ABBREVIATED PROCEDURES OF THE WECHSLER SCALES USING THE "SPLIT-HALF" METHOD

Appendix A contains three examples of the "split-half" method:

(a) An Abbreviated Procedure for the WAIS (Satz & Mogel, 1962)

(b) An Abbreviated Procedure for the WISC (Yudin, 1966)

(c) A Procedure for Reducing the Number of Items in Subtests of the WISC (Silverstein, 1968)

AN ABBREVIATED PROCEDURE FOR THE WAIS
(SATZ & MOGEL, 1962)

Subtest	Items Used	Multiply Score By	Correction
Information	Every 3rd	3	Subtract 1 from scaled score
Comprehension	Odd only	2	
Arithmetic	Odd only	2	
Similarities	Odd only	2	
Digit Span	Unchanged	1	
Vocabulary	Every 3rd	3	
Digit Symbol	Unchanged	1	
Picture Completion	Every 3rd	3	
Block Design	Odd only	2	Subtract 1 from scaled score
Picture Arrangement	Odd only	2	
Object Assembly	Odd only	2	

AN ABBREVIATED PROCEDURE FOR THE WISC
(YUDIN, 1966)

Subtest	Recommended Items	Multiply Score By	Correction Factors
Information	Every 3rd	3	Subtract 1 from scaled score
Comprehension	Odd only	2	None
Arithmetic	Even only	2	Add 1 to scaled score
Similarities	Odd only	2	None
Vocabulary	Every 3rd	3	Subtract 4 from scaled score
Digit Span	All items	1	None
Picture Completion	Every 3rd	3	Subtract 2 from raw score
Picture Arrangement	Odd only	2	Subtract 3 from raw score
Block Design	Odd only	2	Subtract 6 from raw score
Object Assembly	Odd only	2	None
Coding	All items	1	None

A PROCEDURE FOR REDUCING THE NUMBER OF ITEMS IN SUBTESTS OF THE WISC (SILVERSTEIN, 1968)

Subtest	Recommended Items	Multiply Score By
Information	Every 3rd	3
Comprehension	Odd only	2
Arithmetic	Even only	2
Similarities	Odd only	2
Vocabulary	Every 3rd	3
Digit Span	Not Administered	-
Picture Completion	Every 3rd	3
Picture Arrangement	Odd only	2
Block Design	Odd only	2
Object Assembly	Odd only	2
Coding	All items	1

APPENDIX B

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC SHORT FORMS (SILVERSTEIN, 1970)

APPENDIX B

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC SHORT FORMS (SILVERSTEIN, 1970)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
V BD	.856	A V BD	.887	I V PA BD	.904	I A V PA BD	.915
I BD	.836	I V BD	.886	A V PA BD	.903	A S V PA BD	.915
I V	.825	S V BD	.885	A S V BD	.900	I S V PA BD	.913
A V	.822	I V OA	.873	S V PA BD	.900	C A V PA BD	.912
V PA	.822	V PA BD	.873	C A V BD	.897	A S V BD OA	.908
S V	.819	C V BD	.869	I A V BD	.897	I A V BD OA	.908
S BD	.817	I V PA	.867	I S V BD	.896	I S V BD OA	.908
I PA	.816	I C BD	.867	S V BD CO	.892	A S V PC BD	.908
I OA	.811	I S BD	.867	I V PA OA	.892	I C V PA BD	.908
V OA	.811	A V OA	.866	A V PC BD	.891	I A V PA OA	.908

Note 1: Abbreviations: I = Information; C = Comprehension; A = Arithmetic; S = Similarities; V = Vocabulary; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: The formula used to compute these part-whole correlations was suggested by Silverstein (1970). The purpose of this modified formula is to correct for the spuriously high coefficients previously found in part-whole correlations.

APPENDIX C

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS
(SATTLER, 1974)

APPENDIX C

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS (SATTLER, 1974)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
V BD	.906	S V BD	.931	I V C BD	.947	S A V PA OA	.963
I BD	.888	I V BD	.929	S V PA BD	.947	S A V PA BD	.962
S BD	.885	I C BD	.928	I C PC BD	.945	S A V BD OA	.960
C BD	.878	I S BD	.925	S A V OA	.944	I C PC BD CO	.960
V OA	.878	V C BD	.924	I V PA BD	.944	I V PC BD CO	.960
V PC	.868	S C BD	.921	I S C BD	.944	S A C PA OA	.960
S V	.864	S V OA	.919	I C PA BD	.944	I S C PA BD	.960
I S	.860	V PA BD	.919	I S PA BD	.943	I V C PC BD	.959
I PC	.858	A V OA	.919	S V PC BD	.943	A V C BD OA	.959
I V	.857	V PC BD	.919	S V BD OA	.943	A V C PA BD	.958

Note 1: Abbreviations: I = Information; S = Similarities; A = Arithmetic; V = Vocabulary; C = Comprehension; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: The formula used to compute these part-whole correlations was suggested by McNemar (1974).